

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/920,118

DATE: 08/09/2001

TIME: 14:19:19

Input Set : A:\10424003999.txt

Output Set: N:\CRF3\08092001\I920118.raw

3 <110> APPLICANT: Menzel, Rolf
 5 <120> TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIRECTED GENE ASSEMBLY
 7 <130> FILE REFERENCE: 10424-003
 C--> 9 <140> CURRENT APPLICATION NUMBER: US/09/920,118
 C--> 9 <141> CURRENT FILING DATE: 2001-07-31
 9 <150> PRIOR APPLICATION NUMBER: 60/222,134
 10 <151> PRIOR FILING DATE: 2000-07-31
 12 <160> NUMBER OF SEQ ID NOS: 22
 14 <170> SOFTWARE: PatentIn version 3.0
 16 <210> SEQ ID NO: 1
 17 <211> LENGTH: 87
 18 <212> TYPE: DNA
 C--> 19 <213> ORGANISM: Artificial ✓
 21 <220> FEATURE:
 22 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
 24 <400> SEQUENCE: 1
 25 aattcgcgtt taaacttaat taaggtaccc attttttggc agatctagac caaaaaatgg 60
 26 gggcgccgc tccccgggtg gcgcgcc 87
 28 <210> SEQ ID NO: 2
 29 <211> LENGTH: 87
 30 <212> TYPE: DNA
 C--> 31 <213> ORGANISM: Artificial ✓
 33 <220> FEATURE:
 34 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
 36 <400> SEQUENCE: 2
 37 aattggcgcg ccacccgggg agcggccgcc cccatttttt ggtctagatc tgccaaaaaa 60
 38 tgggtacctt aattaagttt aaacgcg 87
 40 <210> SEQ ID NO: 3
 41 <211> LENGTH: 50
 42 <212> TYPE: DNA
 C--> 43 <213> ORGANISM: Artificial ✓
 45 <220> FEATURE:
 46 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
 48 <400> SEQUENCE: 3
 49 gactgcgaga tcatagatat agatttcact acgcggctgc tcaaacctgg 50
 51 <210> SEQ ID NO: 4
 52 <211> LENGTH: 50
 53 <212> TYPE: DNA
 C--> 54 <213> ORGANISM: Artificial ✓
 56 <220> FEATURE:
 57 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
 59 <400> SEQUENCE: 4
 60 ccaggtttga gcagccgcgt agtgaaatct atatctatga tctcgcagtc 50
 62 <210> SEQ ID NO: 5
 63 <211> LENGTH: 87
 64 <212> TYPE: DNA
 C--> 65 <213> ORGANISM: Artificial ✓

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/920,118

DATE: 08/09/2001

TIME: 14:19:19

Input Set : A:\10424003999.txt

Output Set: N:\CRF3\08092001\I920118.raw

67 <220> FEATURE:
68 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
70 <400> SEQUENCE: 5
71 aatttaccat ggagcaattg catatggttt aaacagctcg agtagatctt gcggccgctt 60
72 ggctagcgtc agctgggtac catgcat 87
74 <210> SEQ ID NO: 6
75 <211> LENGTH: 87
76 <212> TYPE: DNA
C--> 77 <213> ORGANISM: Artificial ✓
79 <220> FEATURE:
80 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
82 <400> SEQUENCE: 6
83 cgcgttatgc atggtaccca gctgacgcta gccaaagcggc cgcaagatct actcgagctg 60
84 tttaaaccat atgcaattgc tccatgg 87
86 <210> SEQ ID NO: 7
87 <211> LENGTH: 19
88 <212> TYPE: DNA
C--> 89 <213> ORGANISM: Artificial ✓
91 <220> FEATURE:
92 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
94 <400> SEQUENCE: 7
95 cgcaawcygt tccttaygg 19
97 <210> SEQ ID NO: 8
98 <211> LENGTH: 20
99 <212> TYPE: DNA
C--> 100 <213> ORGANISM: Artificial ✓
102 <220> FEATURE:
103 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
105 <400> SEQUENCE: 8
106 gccaggagcc atsacwtcaa 20
108 <210> SEQ ID NO: 9
109 <211> LENGTH: 28
110 <212> TYPE: DNA
C--> 111 <213> ORGANISM: Artificial ✓
113 <220> FEATURE:
114 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
116 <400> SEQUENCE: 9
117 ggggtaccgc ggtctattca tactttcg 28
119 <210> SEQ ID NO: 10
120 <211> LENGTH: 36
121 <212> TYPE: DNA
C--> 122 <213> ORGANISM: Artificial ✓
124 <220> FEATURE:
125 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
127 <400> SEQUENCE: 10
128 gcagatctca ttgttagaa tatgttattg agcggc 36
130 <210> SEQ ID NO: 11
131 <211> LENGTH: 27
132 <212> TYPE: DNA

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/920,118

DATE: 08/09/2001

TIME: 14:19:19

Input Set : A:\10424003999.txt

Output Set: N:\CRF3\08092001\I920118.raw

C--> 133 <213> ORGANISM: Artificial ✓
 135 <220> FEATURE:
 136 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
 138 <400> SEQUENCE: 11
 139 agcgagatct ctattattgt gcagctg 27
 141 <210> SEQ ID NO: 12
 142 <211> LENGTH: 33
 143 <212> TYPE: DNA

C--> 144 <213> ORGANISM: Artificial ✓
 146 <220> FEATURE:
 147 <223> OTHER INFORMATION: Description of artificial sequence: Primer ✓
 149 <400> SEQUENCE: 12
 150 gcgcggtacc tgataaaagg agagggtaaa gag 33
 152 <210> SEQ ID NO: 13
 153 <211> LENGTH: 1140
 154 <212> TYPE: DNA
 155 <213> ORGANISM: Bacillus licheniformis
 157 <220> FEATURE:
 158 <221> NAME/KEY: CDS
 159 <222> LOCATION: (1)..(1140)
 161 <400> SEQUENCE: 13
 162 atg atg agg aaa aag agt ttt tgg ctt ggg atg ctg acg gcc tta atg 48
 163 Met Met Arg Lys Lys Ser Phe Trp Leu Gly Met Leu Thr Ala Leu Met
 164 1 5 10 15
 165 ctc gtg ttc acg atg gcc ttc agc gat tcc gcg tct gct gct cag ccg 96
 166 Leu Val Phe Thr Met Ala Phe Ser Asp Ser Ala Ser Ala Ala Gln Pro
 167 20 25 30
 168 gcg aaa aat gtt gaa aag gat tat att gtc gga ttt aag tcg gga gtg 144
 169 Ala Lys Asn Val Glu Lys Asp Tyr Ile Val Gly Phe Lys Ser Gly Val
 170 35 40 45
 171 aaa acc gca tcc gtc aaa aag gac atc atc aaa gag agc ggc gga aaa 192
 172 Lys Thr Ala Ser Val Lys Lys Asp Ile Ile Lys Glu Ser Gly Gly Lys
 173 50 55 60
 174 gtg gac aag cag ttt aga atc atc aac gcg gca aaa gcg aag cta gac 240
 175 Val Asp Lys Gln Phe Arg Ile Ile Asn Ala Ala Lys Ala Lys Leu Asp
 176 65 70 75 80
 177 aaa gaa gcg ctt gag gaa gtc aaa aat gat ccg gat gtc gct tat gtg 288
 178 Lys Glu Ala Leu Glu Glu Val Lys Asn Asp Pro Asp Val Ala Tyr Val
 179 85 90 95
 180 gaa gag gat cac gta gct cat gct ttg gcg caa acc gtt cct tac ggc 336
 181 Glu Glu Asp His Val Ala His Ala Leu Ala Gln Thr Val Pro Tyr Gly
 182 100 105 110
 183 att cct ctc att aaa gcg gac aaa gtg cag gct caa ggc tac aag gga 384
 184 Ile Pro Leu Ile Lys Ala Asp Lys Val Gln Ala Gln Gly Tyr Lys Gly
 185 115 120 125
 186 gcg aac gta aaa gtc gcc gtc ctg gat aca gga atc caa gct tct cat 432
 187 Ala Asn Val Lys Val Ala Val Leu Asp Thr Gly Ile Gln Ala Ser His
 188 130 135 140
 189 ccg gac ttg aac gta gtc ggc gga gca agc ttc gta gct ggc gaa gct 480

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/920,118

DATE: 08/09/2001

TIME: 14:19:19

Input Set : A:\10424003999.txt

Output Set: N:\CRF3\08092001\I920118.raw

```

190 Pro Asp Leu Asn Val Val Gly Gly Ala Ser Phe Val Ala Gly Glu Ala
191 145          150          155          160
192 tat aac acc gac ggc aac gga cac ggc acg cat gtt gcc ggt aca gta      528
193 Tyr Asn Thr Asp Gly Asn Gly His Gly Thr His Val Ala Gly Thr Val
194          165          170          175
195 gct gcg ctt gac aat aca acg ggt gta tta ggc gtt gcg ccg aac gta      576
196 Ala Ala Leu Asp Asn Thr Thr Gly Val Leu Gly Val Ala Pro Asn Val
197          180          185          190
198 tcc ttg tac gcg gtt aaa gtg ctg aat tca agc gga agc gga tct tac      624
199 Ser Leu Tyr Ala Val Lys Val Leu Asn Ser Ser Gly Ser Gly Ser Tyr
200          195          200          205
201 agc ggc att gta agc gga atc gag tgg gcg acg aca aac ggc atg gat      672
202 Ser Gly Ile Val Ser Gly Ile Glu Trp Ala Thr Thr Asn Gly Met Asp
203          210          215          220
204 gtt atc aac atg agc ctt gga gga cca tca ggc tca aca gcg atg aaa      720
205 Val Ile Asn Met Ser Leu Gly Gly Pro Ser Gly Ser Thr Ala Met Lys
206 225          230          235          240
207 cag gcg gtt gac aat gca tat gca aga ggg gtt gtc gtt gtg gcg gct      768
208 Gln Ala Val Asp Asn Ala Tyr Ala Arg Gly Val Val Val Val Ala Ala
209          245          250          255
210 gct ggg aac agc gga tct tca gga aac acg aat aca atc ggc tat cct      816
211 Ala Gly Asn Ser Gly Ser Ser Gly Asn Thr Asn Thr Ile Gly Tyr Pro
212          260          265          270
213 gcg aaa tac gac tct gtc atc gca gtt ggc gcg gta gac cct aac agc      864
214 Ala Lys Tyr Asp Ser Val Ile Ala Val Gly Ala Val Asp Pro Asn Ser
215          275          280          285
216 aac aga gct tca ttt tcc agc gtc gga gca gag ctt gaa gtc atg gct      912
217 Asn Arg Ala Ser Phe Ser Ser Val Gly Ala Glu Leu Glu Val Met Ala
218          290          295          300
219 cct ggc gca ggc gtg tac agc act tac cca acc agc act tat gca aca      960
220 Pro Gly Ala Gly Val Tyr Ser Thr Tyr Pro Thr Ser Thr Tyr Ala Thr
221 305          310          315          320
222 ttg aac gga acg tca atg gct tct cct cat gta gcg gga gca gca gct      1008
223 Leu Asn Gly Thr Ser Met Ala Ser Pro His Val Ala Gly Ala Ala Ala
224          325          330          335
225 ttg atc ttg tca aaa cat ccg aac ctt tca gct tca caa gtc cgc aac      1056
226 Leu Ile Leu Ser Lys His Pro Asn Leu Ser Ala Ser Gln Val Arg Asn
227          340          345          350
228 cgt ctc tcc agt acg gcg act tat ttg gga agc tcc ttc tac tat gga      1104
229 Arg Leu Ser Ser Thr Ala Thr Tyr Leu Gly Ser Ser Phe Tyr Tyr Gly
230          355          360          365
231 aaa ggt ctg atc aat gtc gaa gct gcc gct caa taa      1140
232 Lys Gly Leu Ile Asn Val Glu Ala Ala Ala Gln
233          370          375
235 <210> SEQ ID NO: 14
236 <211> LENGTH: 379
237 <212> TYPE: PRT
238 <213> ORGANISM: Bacillus licheniformis
240 <400> SEQUENCE: 14

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/920,118

DATE: 08/09/2001

TIME: 14:19:19

Input Set : A:\10424003999.txt

Output Set: N:\CRF3\08092001\I920118.raw

```

241 Met Met Arg Lys Lys Ser Phe Trp Leu Gly Met Leu Thr Ala Leu Met
242 1 5 10 15
244 Leu Val Phe Thr Met Ala Phe Ser Asp Ser Ala Ser Ala Ala Gln Pro
245 20 25 30
247 Ala Lys Asn Val Glu Lys Asp Tyr Ile Val Gly Phe Lys Ser Gly Val
248 35 40 45
250 Lys Thr Ala Ser Val Lys Lys Asp Ile Ile Lys Glu Ser Gly Gly Lys
251 50 55 60
253 Val Asp Lys Gln Phe Arg Ile Ile Asn Ala Ala Lys Ala Lys Leu Asp
254 65 70 75 80
256 Lys Glu Ala Leu Glu Glu Val Lys Asn Asp Pro Asp Val Ala Tyr Val
257 85 90 95
259 Glu Glu Asp His Val Ala His Ala Leu Ala Gln Thr Val Pro Tyr Gly
260 100 105 110
262 Ile Pro Leu Ile Lys Ala Asp Lys Val Gln Ala Gln Gly Tyr Lys Gly
263 115 120 125
265 Ala Asn Val Lys Val Ala Val Leu Asp Thr Gly Ile Gln Ala Ser His
266 130 135 140
268 Pro Asp Leu Asn Val Val Gly Gly Ala Ser Phe Val Ala Gly Glu Ala
269 145 150 155 160
271 Tyr Asn Thr Asp Gly Asn Gly His Gly Thr His Val Ala Gly Thr Val
272 165 170 175
274 Ala Ala Leu Asp Asn Thr Thr Gly Val Leu Gly Val Ala Pro Asn Val
275 180 185 190
277 Ser Leu Tyr Ala Val Lys Val Leu Asn Ser Ser Gly Ser Gly Ser Tyr
278 195 200 205
280 Ser Gly Ile Val Ser Gly Ile Glu Trp Ala Thr Thr Asn Gly Met Asp
281 210 215 220
283 Val Ile Asn Met Ser Leu Gly Gly Pro Ser Gly Ser Thr Ala Met Lys
284 225 230 235 240
286 Gln Ala Val Asp Asn Ala Tyr Ala Arg Gly Val Val Val Val Ala Ala
287 245 250 255
289 Ala Gly Asn Ser Gly Ser Ser Gly Asn Thr Asn Thr Ile Gly Tyr Pro
290 260 265 270
292 Ala Lys Tyr Asp Ser Val Ile Ala Val Gly Ala Val Asp Pro Asn Ser
293 275 280 285
295 Asn Arg Ala Ser Phe Ser Ser Val Gly Ala Glu Leu Glu Val Met Ala
296 290 295 300
298 Pro Gly Ala Gly Val Tyr Ser Thr Tyr Pro Thr Ser Thr Tyr Ala Thr
299 305 310 315 320
301 Leu Asn Gly Thr Ser Met Ala Ser Pro His Val Ala Gly Ala Ala Ala
302 325 330 335
304 Leu Ile Leu Ser Lys His Pro Asn Leu Ser Ala Ser Gln Val Arg Asn
305 340 345 350
307 Arg Leu Ser Ser Thr Ala Thr Tyr Leu Gly Ser Ser Phe Tyr Tyr Gly
308 355 360 365
310 Lys Gly Leu Ile Asn Val Glu Ala Ala Ala Gln
311 370 375
313 <210> SEQ ID NO: 15

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/920,118

DATE: 08/09/2001

TIME: 14:19:20

Input Set : A:\10424003999.txt

Output Set: N:\CRF3\08092001\I920118.raw

L:9 M:270 C: Current Application Number differs, Replaced Current Application No
L:9 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:19 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:1
L:31 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:2
L:43 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:3
L:54 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:4
L:65 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:5
L:77 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:6
L:89 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:7
L:100 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:8
L:111 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:9
L:122 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:10
L:133 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:11
L:144 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:12
L:477 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:17
L:488 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:18